CLAIMS

What is claimed is:

1. A method for generating an electronically publishable document, comprising:

receiving image data corresponding to a physical document;

segmenting said image data;

creating a markup language file containing said segmented image data; and

embedding a graphical user interface within said markup language file that enables navigation to segmented image data as selected by the user.

2. The method of claim 1 further comprising:

performing optical character recognition (OCR) processing of the segmented imaged data.

3. The method of claim 2 further comprising:

searching text data generated from said OCR processing to identify section identifiers.

4. The method of claim 3 further comprising:

creating a plurality of links in said markup language file utilizing said section identifiers to enable user navigation to said segmented image data associated with said section identifiers.

5. The method of claim 4 wherein said plurality of links are created in a table of contents section of said markup language file.

6. The method of claim 2 wherein said embedding a graphical user interface comprises:

embedding a script in said markup language file that performs a search of document text in response to search queries.

25260034.1

7.   The method of claim 1 wherein said physical document is a multi-page document, said method further comprising:

creating page identifiers within said markup language file.

8.   The method of claim 7 wherein said embedding a graphical user interface comprises:

providing user controls to enable user navigation according to said page identifiers.

9.   The method of claim 1 wherein said embedding a graphical user interface comprises:

receiving manual identification of ones of said segmented image data; and

creating links within said markup language file to enable user navigation to said manually identified ones of said segmented image data.

10.   A computer readable medium containing executable instructions for generating an electronically publishable document, said computer readable medium comprising:

code for segmenting image data of a physical document;

code for creating a markup language file;

code for encapsulating said segmented image data within said markup language file; and

code for embedding a graphical user interface within said markup language file that enables navigation to said segmented image data in response to user input.

11.   The computer readable medium of claim 10 further comprising:

code for generating a text file from image data segmented by said code for segmenting.

12. The computer readable medium of claim 11 further comprising:

code for creating a search control within said markup language file to enable user navigation according to text queries.

13. The computer readable medium of claim 11 further comprising:

code for searching said text file to identify keywords indicative of a section of said physical document; and

code for creating links in said markup language document to enable user navigation to segmented image data corresponding to keywords identified by said code for searching.

14. The computer readable medium of claim 10 further comprising:

code for creating markup language tags in said markup language file to indicate segmented image data corresponding to pages of said physical document.

15. The computer readable medium of claim 14 further comprising:

code for embedding a user control in said markup language file to enable user navigation to a selected page.

16. The computer readable medium of claim 14 further comprising:

code for embedding a user control in said markup language file to enable user navigation forward or backward according to said markup language tags that indicate segmented image data corresponding to pages of said physical document.

17. The computer readable medium of claim 10, wherein said code for segmenting segments image data corresponding to text elements, line art elements, and graphical image elements.

18.   A system for generating an electronically publishable document, comprising:

means for providing image data;

means for performing page segmentation on said image data;

means for creating a markup language file containing segmented data generated by said means for performing page segmentation; and

means for embedding a graphical user interface within said markup language file to enable navigation to said segmented data according to user input.

19.   The system of claim 18 further comprising:

means for performing optical character recognition (OCR) processing upon text data segmented by said means for performing page segmentation.

20.   The system of claim 19 further comprising:

means for embedding a search script, in said markup language file, operable to search text data generated by said means for performing OCR processing to enable navigation to ones of said segmented data according to search queries.